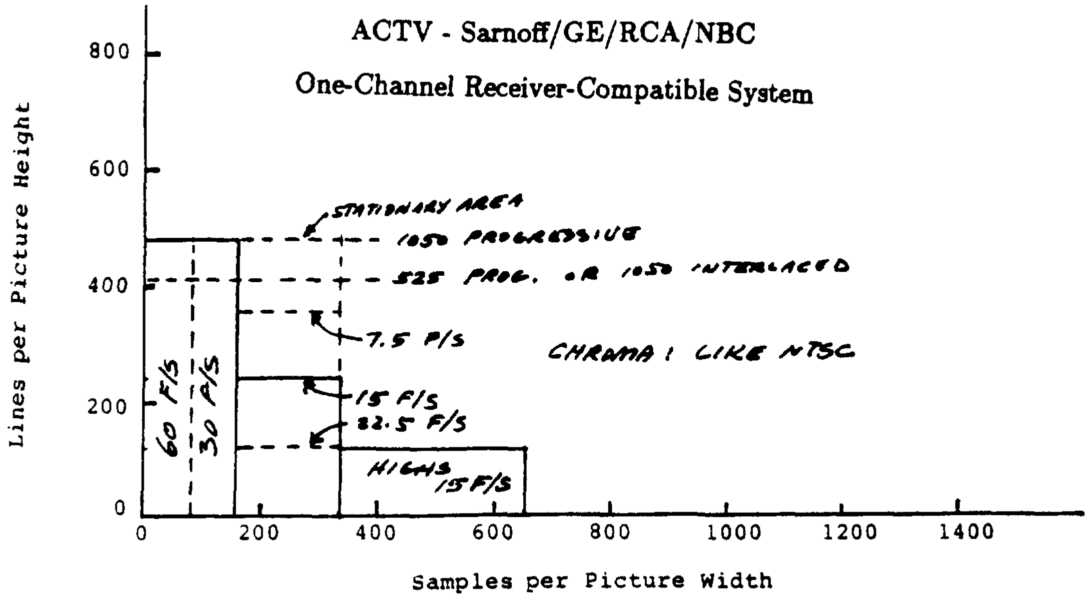



### *Isnardi "ACTV" (NBC/RCA/Sarnoff) System*

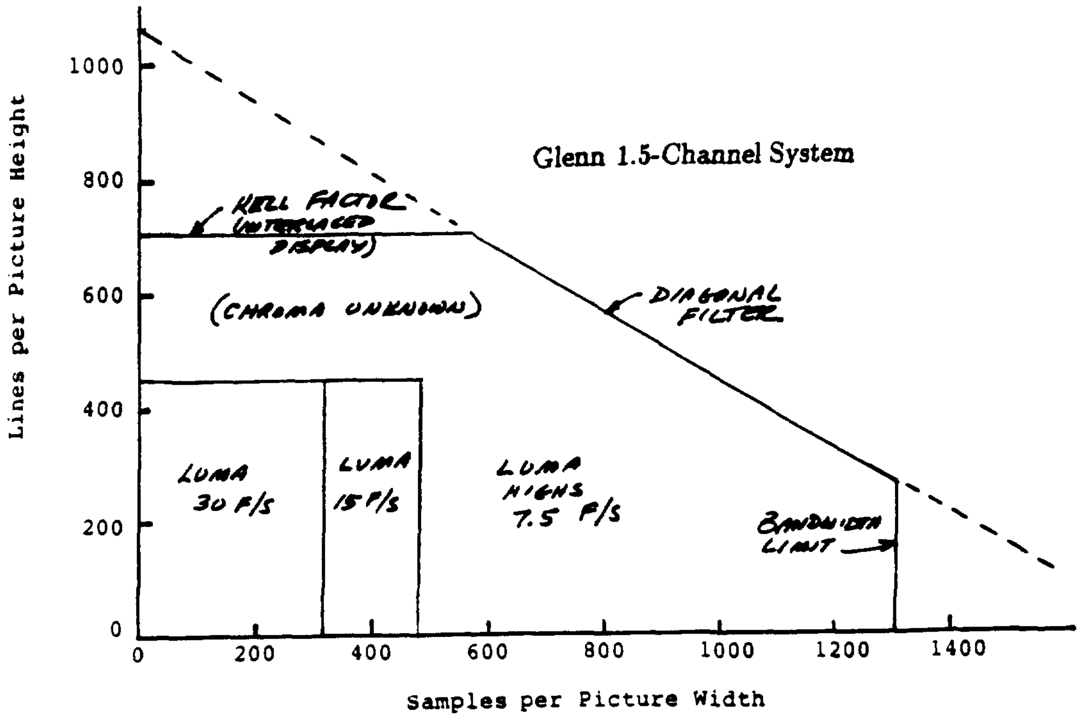
This uses both the Fukinuki and Yasumoto techniques to add 3 low-level enhancement signals for increasing spatial resolution and decreasing the effects of interlace. The low-frequency component of the side panels is time-compressed and placed at the edges of the NTSC image, hidden by the picture frame. All the other signals are highs about 12 db down from the main signal. The receiver can be 525 lines progressive, 1050 lines interlaced, or 1050 lines progressive. There is no enhancement of chroma resolution or of audio. This system has been simulated but not tested under realistic conditions. In all likelihood, some modifications will be required for operation over noisy channels with multipath.

### *Yasumoto (Matsushita) System (not shown)*

An extra channel is provided in the NTSC format by adding a second carrier in quadrature with the main carrier. The two signals can be separated properly in receivers using synchronous detectors, but crosstalk occurs in envelope-detector receivers. This requires that the extra signal, which may have a bandwidth of about 1 MHz, be of rather small amplitude. The scheme permits about a 25% increase in horizontal resolution, but probably cannot be used to add side panels. This system has been demonstrated.

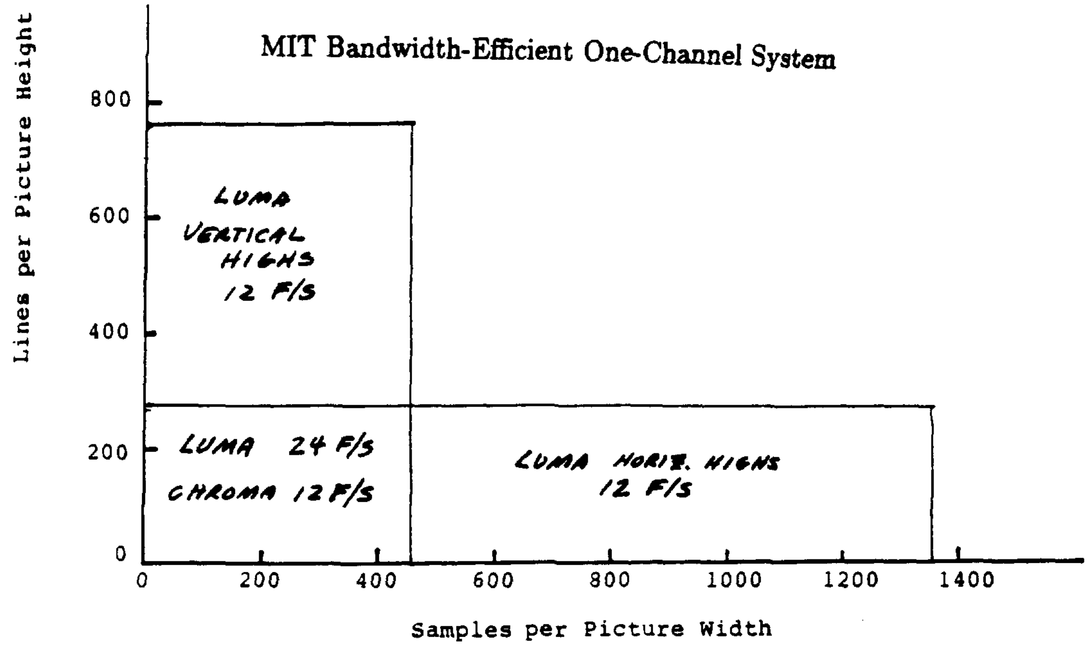
### *Fukinuki (Hitachi) System (not shown)*

This scheme uses a second subcarrier, similar to the color subcarrier, permitting about 1.2 MHz of additional signal of reduced amplitude. (The Sarnoff system uses such a carrier to add more than 2 MHz.) This is done at the expense of diagonal luminance resolution of moving objects and of vertical chroma resolution. Three-dimensional filtering with frame stores is required at both transmitter and receiver. Because of increased cross color and cross luminance, the enhancement signals must be of low amplitude and thus cannot be used to add side panels. This system has been demonstrated.



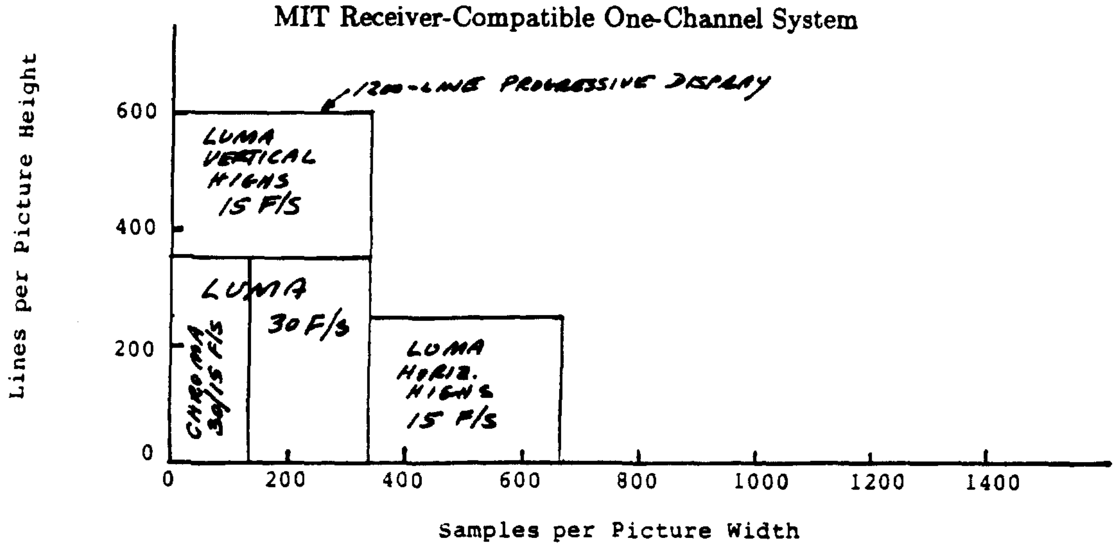

*Glenn System*

**This is similar to the Philips System except that the augmentation signal is a low-frame-rate version of the spatial high-frequency component of an HD signal with exactly twice NTSC resolution. The highs signal, which is transmitted in a full or half channel, (or, in an earlier version, in 2.05 MHz) is found by taking the difference between the HD signal and the NTSC signal, the latter being obtained from the former by filtering and subsampling. (That is the way I would do it - Glenn actually uses two separate cameras.) Compression of the highs information is by diagonal subampling as in MUSE and by using 15 or 7.5 frames/sec. The 7.5 fps, 1.5-channel version is shown. The spatial resolution in stationary areas is the full HD image, but newly-revealed areas take two or four frames to reach their full sharpness (this causes no problems) and moving detail is somewhat blurred. Glenn claims that this effect can be reduced by spatiotemporal filtering of the NTSC signal. A wider aspect ratio, but less than 5:3, is achieved in the enhanced image by removing part of the height of the NTSC image and also intruding into the horizontal blanking time. The system has been demonstrated, but picture quality was reduced due to imperfect hardware.**




*MIT-BE Bandwidth-Efficient System*

This is a family of 6-MHz single-channel HDTV system. Part of the improvement comes from increasing the efficiency of channel utilization by using double-sideband quadrature modulation and elimination of the sound carrier and retrace intervals. A frame store and sophisticated signal processing is required in the receiver. Provision is made for some data transmission as well as digital audio. The signal is divided into components, with 24 fps used for low spatial frequency RGB, and with 12 fps transmission of luminance horizontal and vertical spatial highs and luminance temporal highs




*MIT-RC Receiver-Compatible System*

A portion of the height of the NTSC frame is used for enhancement signals, leaving a 16:9 unimpaired area for NTSC receivers. The enhancement signals are used in a special receiver to raise the vertical resolution to 600 lph and the horizontal resolution to 660 ppw. There is no enhancement of diagonal resolution. It is thought that digital audio might be added into the enhancement signals, which can otherwise be transmitted at full amplitude. Horizontal chroma resolution enhancement is planned by subsampling.

References


1. W.F.Schreiber, A.B.Lippman, A.N.Netravali, and E.D.Adelson, "A Single-Channel Noncompatible High-Definition HDTV System," ATRP-T-63, May 26, 1987, unpublished. W.F.Schreiber, "6-MHz Single-Channel HDTV Systems," 3rd HDTV Symposium, Ottawa, Oct. 87.

2. E.Dubois and P.Faubert, "Two-Dimensional Filters for NTSC Encoding and Decoding," IBC 86.

3. B.Wendland and H.Schroeder, "Signal Processing for New HQTV Systems," SMPTE J, 84, pp 182-189, Oct. 84.

4. W.E.Glenn, "Reduced Bandwidth Requirements for Compatible HDTV System," presented at IGC HDTV Conference, New Orleans, April, 84.

5. Philips

6. T.S.Rzeszewski, "A Compatible High-Definition Television System," Bell System Tech. J, 62, 7, Sept. 83, pp 2091-2111.

7. T.Fukinuki and H.Hirano, "Extended Definition TV Fully Compatible with Existing Standards," IEEE Trans. Comm., 32, pp 948-953, Aug. 84.

8. Y.Yasumoto et al, "An Extended Definition Television System Using Quadrature Modulation of the Video Carrier with Inverse Nyquist Filter," Consumer Electronics Conference, Chicago, 1987.

9. M. Isnardi et al, "A Single Channel, NTSC Compatible Widescreen EDTV System," 3rd HDTV Symposium, Ottawa, Oct. 87. M. Isnardi, "Exploring and Exploiting Subchannels in the NTSC Spectrum," presented at SMPTE Conference, Los Angeles, Oct. 87.

10. R.J.Iredale, "A Proposal for a New High Definition NTSC Broadcast Protocol," IEEE Trans. Consumer Electronics, CE-33, 1, pp 14-27, Feb. 87.

11. Y.Ninomiya et al, "A Single Channel HDTV Broadcast System - the Muse." NHK Laboratories Note No. 304, September 1984.

12. W.F.Schreiber, "Improved Television Systems: NTSC and Beyond," SMPTE J, 66,8, pp 734-743, Aug. 87.

13. P.T.King, "A Novel Television Add-On Data Communication System," SMPTE J, 83,1, pp 10-13, Jan 74.

14. T.Fujio et al, "HDTV," NHK Tech. Monograph No. 32, June 1982.

15. W.F.Schreiber and R.R.Buckley, "A Two-Channel Picture Coding System: II - Adaptive Companding and Color Coding, IEEE Trans. on Communications, COM-29, 12, pp 1849-1858, Dec. 81.

# 6-MHZ SINGLE-CHANNEL HDTV SYSTEMS

William F. Schreiber
Massachusetts Institute of Technology

*Abstract*

After a short description of our TV research program at MIT, considerations related to the character and manner of introduction of advanced television systems are presented. The importance and feasibility of using a modern, sophisticated, programmable receiver in new television systems are stressed. With such a receiver, the system can be improved over time without obsolescence, for example by introducing motion-compensated temporal interpolation. Given such a receiver, a number of techniques are described for more efficient use of the analog 6-MHz channel. It is also proposed to resolve the signal into a number of components and to assign channel capacity to each according to psychophysical principles. On this basis, it is believed that very high quality transmission would be possible in one channel. A backward-compatible version of the system is presented, which is thought to give nearly Muse quality, also in one channel, by means of using a portion of the vertical height of the NTSC frame for enhancement information.

Please note that this is an informal paper. No references are given. Many of the ideas included are not novel, and/or are not those of the author. My colleagues at MIT have made many contributions to this work. A more complete exposition of the two systems will be presented at the SMPTE TV Conference in Nashville next January.

## The Advanced Television Research Program

ATRP is an MIT research program funded by the members of the Center for Advanced Television Studies (CATS). Each member has a 3-year contract with MIT, paying $100,000 per year. The Public Broadcasting System (PBS), which does not make a monetary contribution, acts as a secretariat for CATS. The members are US TV broadcasters and other American companies interested in the TV industry. The program has Justice Department approval under the antitrust laws. The first contract ran from 1983. The original sponsors were ABC, NBC, Time, Inc. (HBO, ATC), PBS, Ampex, Tektronix, RCA, Harris, 3M, and CBS, of which all but the last three renewed in 1986. Zenith and Kodak are current members. CATS is seeking additional members.

The main purpose of the program is to carry out faculty/student research to learn how to make better TV systems, either by improvement of NTSC or by development of entirely new systems. The research agenda is set by mutual agreement. Provision is made for sponsor personnel to spend time at MIT. The program also provides opportunities for the sponsors to meet, free of antitrust restrictions, to discuss matters of mutual interest related to the objectives.

Much of the work is by computer simulation. We have a VAX 11/785 for our own use and have access to many other computer systems, including a Connection Machine. The HDTV simulation facility includes several 1" VTR's, a high-speed disk system, a real-time programmable 3-d *(x,y,t)* interpolator for line- and frame-rate upconversion, 1000-line and 2000-line progressively scanned displays, and a frame-at-a-time 70-mm motion-picture film scanner, all interfaced to the computer. Simulated output sequences are stored on the disk and then played out in real time, upconverted, and viewed on one of the monitors. At present we are limited to a data rate of 8 MB/s into the interpolator, but, by the end of the year, this will be raised to about 72 MB/s. NTSC output is produced by recording sequences, frame by frame, on the VTR and then viewing them at normal speed. We have also simulated a number of advanced TV systems proposed by others.

We have built an Audience Research Facility in a large suburban shopping mall. Subjects are recruited with a small gift, after which they view TV and/or listen to audio, answering questions and filling out questionnaires about their reaction to the material. We have carried out a number of studies on picture and sound quality, and expect to compare 1125-line HDTV with NTSC this fall.

Research Agenda

During the first several years, we assiduously avoided trying to design new TV systems, concentrating at first on methods, NTSC improvement, and fundamental studies of the image-transmission process. We have studied filtering, sampling, and interpolation of moving picture signals, and created a computer model of the entire process. We are actively studying means to control cross effects in NTSC. We have done a good deal of work in motion estimation, and applied it to motion-compensated temporal interpolation and noise reduction. For example, we have been able to get rather good motion rendition in NTSC from simulated 12 frame/sec motion pictures, and have successfully averaged over about 6 frames of a sequence for noise reduction with only negligible blurring. As a demonstration of the effectiveness of adaptive interpolation, we have shown that it is possible to lengthen or shorten a TV sequence by as much as 20% with very high quality. A project on restoration of motion-blurred images is underway, as well as a number of efforts in data compression. We have also simulated a novel adaptive FM system, in which the signal is divided into highs and lows, and the highs are subject to a variable modulation index, thus reducing noise in blank areas where it otherwise would be most noticeable. By this means, we believe it would be possible to get about 12 MHz basebandwidth in a 24-MHz satellite channel.

More recently, we have been studying the design of HDTV systems that are economical of bandwidth. Two such systems are described below. Simulations of both systems are now under consideration under our current sponsorship.

The Future of Television

The subject that has dominated most of our discussions with sponsors is not only what should be the nature of improved TV systems, but how they can be introduced in an acceptable way - acceptable to the public and to the TV industry - and be of maximum benefit to the country as a whole. These discussions have been very instructive, particularly on the question of the effect of new technologies on the audience share of the various TV distribution entities. One thing that is very clear is that the approximately 140 million receivers now in use in the US *must* continue to be served. Most industry people interpret this principle to mean that any new system must be compatible, and they point to the NTSC color method as a model. I have pointed out that PAL was introduced in the UK, and SECAM in France, in a noncompatible manner, the existing receivers being served by separate broadcasts.[1] This argument has not been effective. On the other hand, the near-unanimous emphasis on *receiver compatibility* among our sponsors, who, I believe, are representative of the TV industry as a whole, seems in conflict with the participation of many individuals and companies in the deliberations of ATSC and SMPTE which, for the most part, are directed to ironing out details of the 1125-line noncompatible system.

There are two drawbacks, in my mind, to single-channel backward-compatible systems, usually called EDTV. One is that the image quality that can be obtained on the special receivers, while at the same time not degrading the quality on existing receivers, is not high enough to qualify as HDTV. (In fact, many people do not even regard Muse as "true" HDTV.) Even if it were, I do not believe that a large share of the viewers would spend substantial extra money solely to receive the same programs in higher technical quality. The second objection is that many of the methods proposed are technological dead ends. There is no way to extend them to the 1000-line or more resolution nearly everyone seems to want, and they get more and more "hairy" as we try to push to higher and higher quality.[2]

Another group of proposals uses an NTSC channel, for compatibility, plus an enhancement channel; the special receiver puts both together to make HDTV. I have no doubt that this method can be made to work. However, it permanently ties up the extra channel, and enshrines the inefficiencies and defects of NTSC in the "new" system. It prevents the development of 6-MHz HDTV systems, which I believe is possible.

If existing receivers can continue to be serviced by NTSC or enhanced NTSC broadcasts, then it may be economically feasible to develop a totally new system, with a goal of very high quality in one channel. Of course, higher quality, in this or any other system, can readily be achieved with more bandwidth. However, bandwidth is *never* over-abundant[3] and what there is, is firmly channelized into 6-MHz chunks, even on cable.

There is good reason to believe that, starting from scratch, and without the constraint of NTSC receiver compatibility, a much better system than NTSC could be developed. It must be remembered that NTSC, in addition to being compatible with its monochrome predecessor, had to assume a very simple receiver. With today's technology, let alone what we can reasonably expect in the 90's, for the same real price as the first color receivers, we can now have something much more sophisticated. For example, it would have been inconceivable to call for a frame store 35 years ago; it is inconceivable that the TV receiver of the future should be without one!

---

[1] Admittedly, conditions were quite different in Europe at that time from in the US today. In might well be *easier* to do this under present conditions, especially if station-ownership rules were changed to accomodate HDTV.

[2] Lesser improvements in NTSC, such as removing cross effects, and possibly the use of a progressive-scan display, may well go ahead under normal market pressures, for their own sake, and not as a stepping stone to HDTV. Some often-discussed improvements would not be so easy to implement, however. For example, elimination of cross color and cross luminance requires prefiltering at the transmitter, an operation completely impractical in the composite studio. In the case of upconverted receivers, very little improvement in displayed vertical resolution, which must originate in the camera, is possible without producing totally unacceptable interline flicker on interlaced receivers.

[3] Witness the alarm with which the TV industry has reacted to the proposal to give part of the UHF band to land-mobile radio.

In the following section, I shall describe some techniques that might be applied to the solution of this problem. They involve a sophisticated receiver, a signal divided into components that can be processed by such a receiver, and a more efficient analog modulation method. Properly configured, such a receiver can also serve in an appropriately designed backward-compatible 6-MHz system. If this is true, then such systems, which are overwhelmingly preferred by broadcasters, need *not* be technological dead ends. Instead, they can be stepping stones to the future, in which, as visualized in the recent FCC Notice of Inquiry, NTSC can eventually be phased out.

Principles of More Efficient TV Systems

*Better analog channel utilization.* Eliminate the separate sound carrier, the retrace intervals, and vestigial sideband transmission. Instead use two 3-MHz baseband signals, quadrature modulated onto a single carrier in the middle of the band. Use time multiplexing of the various components, reserving about 1/12 of the time for audio and data. This gives an improvement of as much as 70% in efficiency (pels/sec per unit bandwidth.)

*Separation of camera and display standards from those of the channel.* This involves a frame store and high-rate progressively scanned display. It gives a very high Kell factor and permits high vertical resolution in the camera without interline flicker in the display. Note that with an interlaced display, the maximum permissible camera vertical resolution without producing such flicker is barely half the theoretical value. This method effectively increases the resolution at least 40%.

*Increased spatial resolution/reduced frame rate.* There is a good deal of evidence that today's 60 fps, which was chosen primarily on account of large-area flicker considerations[4] is a poor tradeoff when a flickerless display, driven by a frame store, is used. The acceptance by viewers of motion rendition in film at 24 fps, and, even worse, when 24-fps films are shown on NTSC by any of the usual methods, indicates that it would be better to use the capacity for higher spatial resolution. After all, that is the principal claim of HDTV. I would guess that 18 frames and 36 fields per second would be entirely satisfactory with a carefully designed nonadaptive interpolator. This should give much better motion than film, and would result in another 2/3 improvement in resolution.

*Reduced relative diagonal resolution.* I am less sure that this technique really works, as our audience tests indicate no special preference for the diamond-shaped passband. However, even making the diagonal resolution equal to the vertical and horizontal resolution gives an improvement of perhaps 20%.

*Relatively lower frame rate for high spatial frequencies.* This technique is used in Muse and by Glenn, as well as in the Philips and Sarnoff/NBC systems. The savings depend on just how the band is divided, but come to at least 25%.

*Relatively lower frame rate for chrominance information.* The applicability of this technique is indicated by psychophysical data indicating that the bandwidth for chrominance temporal variations is considerably lower than for luminance. Since we propose to use chrominance

[4] I suspect the power line frequency itself was chosen for the same reasons!

spatial resolution about 1/3 that of luminance in all directions (rather than the very unequal resolution used in NTSC), only about a 5% improvement would come from this technique. However, it might be useful in a mixed-highs system, which we favor over a luminance/chrominance system, from SNR considerations.

If all these improvements in efficiency multiply, then we shall have about a 6-fold increase in the area spatial resolution compared to NTSC, even with nonadaptive interpolation. This would give a quality at least equal to Muse, with the possibility of eventually getting much better.

*Improvement in signal-to-noise ratio.* To my knowledge, no one has proposed that higher quality TV demands higher SNR on the pictures actually seen in the home. However, it is quite clear that, at least in the US, the link between transmitter and receiver is indeed the weakest link in NTSC. I rather suspect that most viewers, shown pictures on a studio monitor, would call that HDTV. In the mixed-highs system, adaptive modulation, similar to that mentioned above in connection with the adaptive FM system, can reduce noise to a very considerable degree in the blank areas, where it is most visible. The degree of improvement depends on how the signal is split into lows and highs. Especially where a data channel is available for transmission of adaptation information, this is a very effective method of improving picture quality.

The Open-Architecture Smart Receiver

The previous section implies that the receiver not only has a frame store, but is capable of separate interpolation (upconversion) of the various components, some adaptively modulated, as well as the assembly of the various components into appropriate video signals for display. If we indulge in some "newthink," that is, look around at the proliferation, in our homes and offices, of very inexpensive but nevertheless very powerful computational and storage components, it becomes reasonable to think about a much more powerful receiver that would probably cost very little extra. In the millions, which is the level we should think about for receivers, computational capability becomes very cheap indeed.

An example of a capability made possible by computational power is flexible transmission frame rate. The receiver can be capable of adapting itself, under the control of a small amount of digital information transmitted along with the picture signal, to the signal format. By this I mean, not only NTSC, SECAM, and PAL[5] (there are already receivers and VCR's that do that), but different frame rates. These can be chosen at the transmitter according to the subject matter, on a program-by-program, scene-by-scene, or even frame-by-frame basis. Sports scenes can be shot at high temporal resolution, and slow-moving outdoor scenes at high spatial resolution. In both cases, the display, as distinct from the transmission, would operate at both high resolution and high frame rate to give a flicker-free picture without visible line structure.

Another set of capabilities can be introduced, at modest cost, by adopting a bus-structured open architecture. If we think of a modular receiver consisting of a tuner, computation section, and display module - the latter including the refresh memory - it is clear that,

[5] It would also be pratical to adapt such a receiver to any of the recently proposed HDTV or EDTV formats.

with the right arrangement, the computation section can be made very flexible, in the same manner as as IBM PC. A programmable receiver of this type would permit improvements in the system to be added at a later date, as we learn more about the video processing problem. It would also allow third parties to design plug-in software or hardware modules to give special capabilities, such as interactivity and easy connection to a wide variety of peripherals. These might include computers, cable, VCR's, video games, and, perhaps most important of all, electronic still photographic equipment.

An exceedingly important capability that might be added at a later date is motion-compensated temporal interpolation. The work at ATRP and elsewhere shows that good motion rendition, i.e., smooth and sharp rendition of moving objects, is possible using motion compensation. At the present time, this is too computationally intensive for serious consideration in receivers. However, if there are enough receivers ready to plug them in, it may well be possible to develop integrated circuits cheap enough to use for this purpose. In that case, much better performance can be achieved from the same receivers, simply by a different trade-off of spatial and temporal resolution at the transmitter.

We have discussed this general idea at MIT for the last six months or so. The more it is discussed, the more I have come to believe that a receiver of this type, with its built-in protection against obsolescence, may well be the single most important ingredient of the TV system of the future.

A Backward-Compatible Single-Channel System

The previous discussion was centered on means for getting very high quality in a single channel in a noncompatible manner. However, when considering the capabilities that are reasonable to incorporate in the receiver, it becomes evident that there is a rather simple way to modify NTSC to achieve very high quality on new receivers, together with acceptable operation on existing receivers.

Iredale and others have suggested that the different aspect ratio of the proposed HDTV systems and of NTSC be accommodated by removing part of the height of the NTSC picture. Others have suggested adding side panels to NTSC, which amounts to the same thing, from the artistic point of view.[6] There is a twist to this that I think is worth considering. If 25% of the NTSC picture height is usurped (60 lines each at top and bottom), we are left with a picture 360 lines high with an aspect ratio of 16:9. If these sections are used for enhancement signals of the kind discussed above, it would be possible to achieve nearly Muse quality on a special receiver. This special receiver could be exactly the same smart receiver previously described. In this application, it would receive NTSC 75% of the time and HDTV-type highs components 25% of the time. Something would have to be done for horizontal chrominance resolution (vertical is already high enough), but that also seems possible.

The standard receiver would see bars of moving noise - the enhancement signals - in these

[6] My own opinion is that both schemes are defective in that they completely ignore the intention of the artist in making the film. For example, Woody Allen thinks aspect ratio is so important that he *forbids* any change when his films are shown on television!

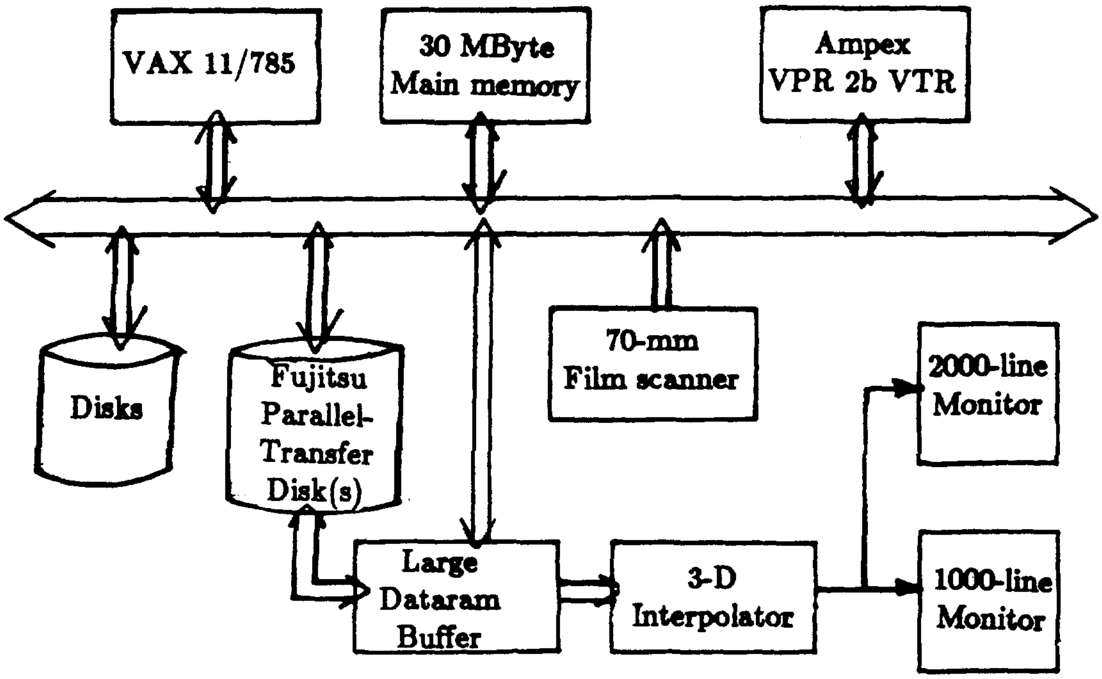
areas. The annoyance level could be reduced somewhat by proper signal design.[7] The usual overscan setting on most receivers would make the bars smaller. If desired, an inexpensive adaptor could blank them out. This adaptor would be especially cheap in receivers connected to cable boxes or VCR's since it would only have to work on one channel. New NTSC receivers could have the blanking arrangement built in at negligible cost. Note that in all cases, the standard receiver would see the entire picture, since the aspect ratio of both formats is exactly the same.

## Conclusion

We have described a set of techniques by which the efficiency of the analog 6-MHz channel could be substantially increased. This, together with a division of the video signal into a set of components, and the assignment of channel capacity to the components in an optimum way, could permit the transmission of very high quality images in a single channel. A smart receiver, required to reassemble the signal and display the image at high line and frame rate, could easily be adapted to a scheme for backward-compatible transmission. In that system, existing receivers would display an image that is unimpaired for the central 360 lines. The smart receiver would also be capable of being upgraded over time so that the system could continue to be improved in quality without obsolescence.

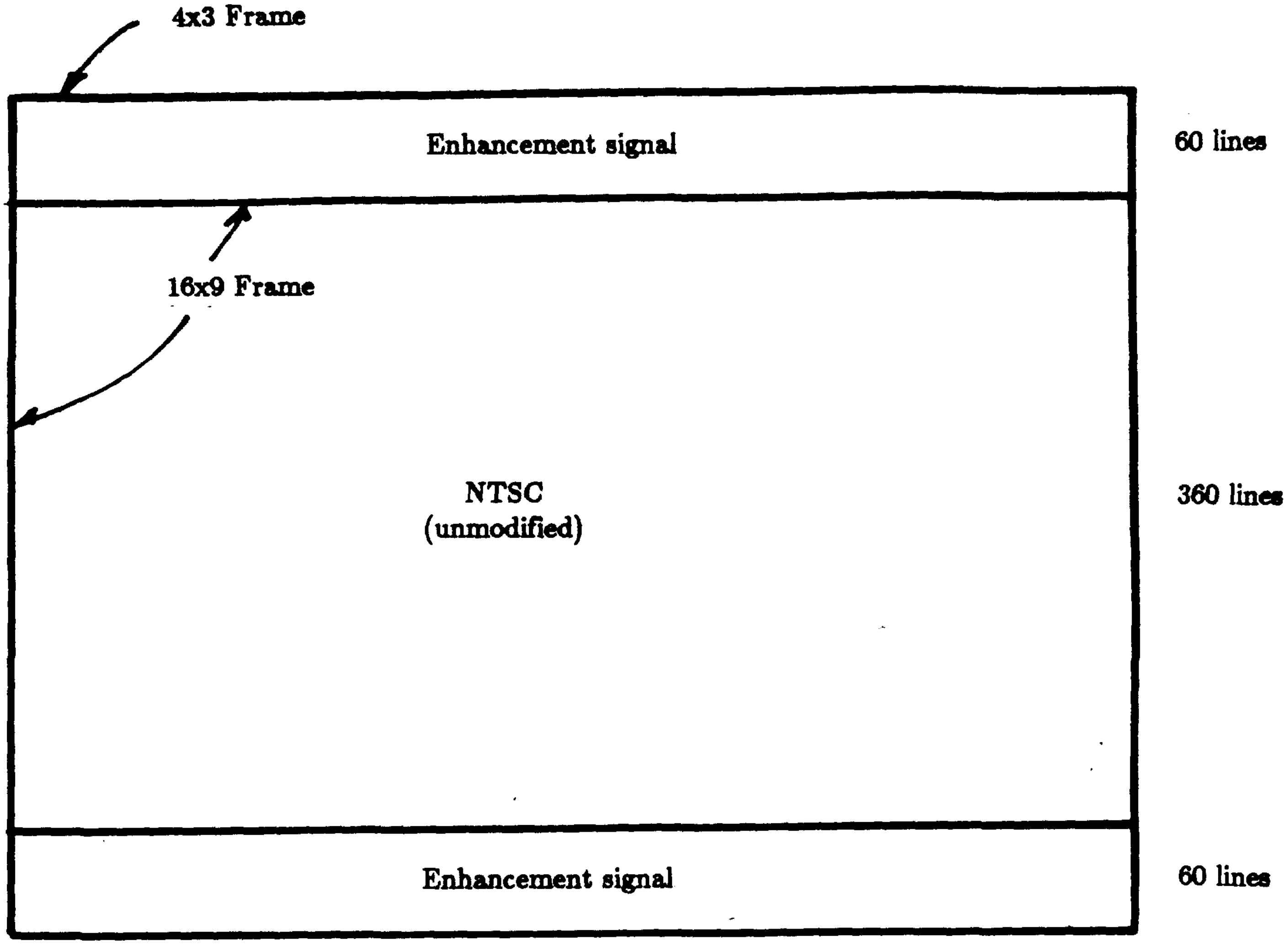
The opinions expressed are those of the author alone, and not of MIT or the members of the Center for Advanced Television Studies. The contribution of students and colleagues to these ideas is gratefully acknowledged.

[7]Although I certainly do not wish to imply that there is an *advantage* in having bars at top and bottom (although it does eliminate the problem of aspect-ratio conversion), making them look like fine-grain noise would considerably reduce their visibility. Like the NBC Peacock of the early days of color ("The following program is brought to you in living color!"), it might serve to remind the viewers that they could see a better picture by buying a better receiver.



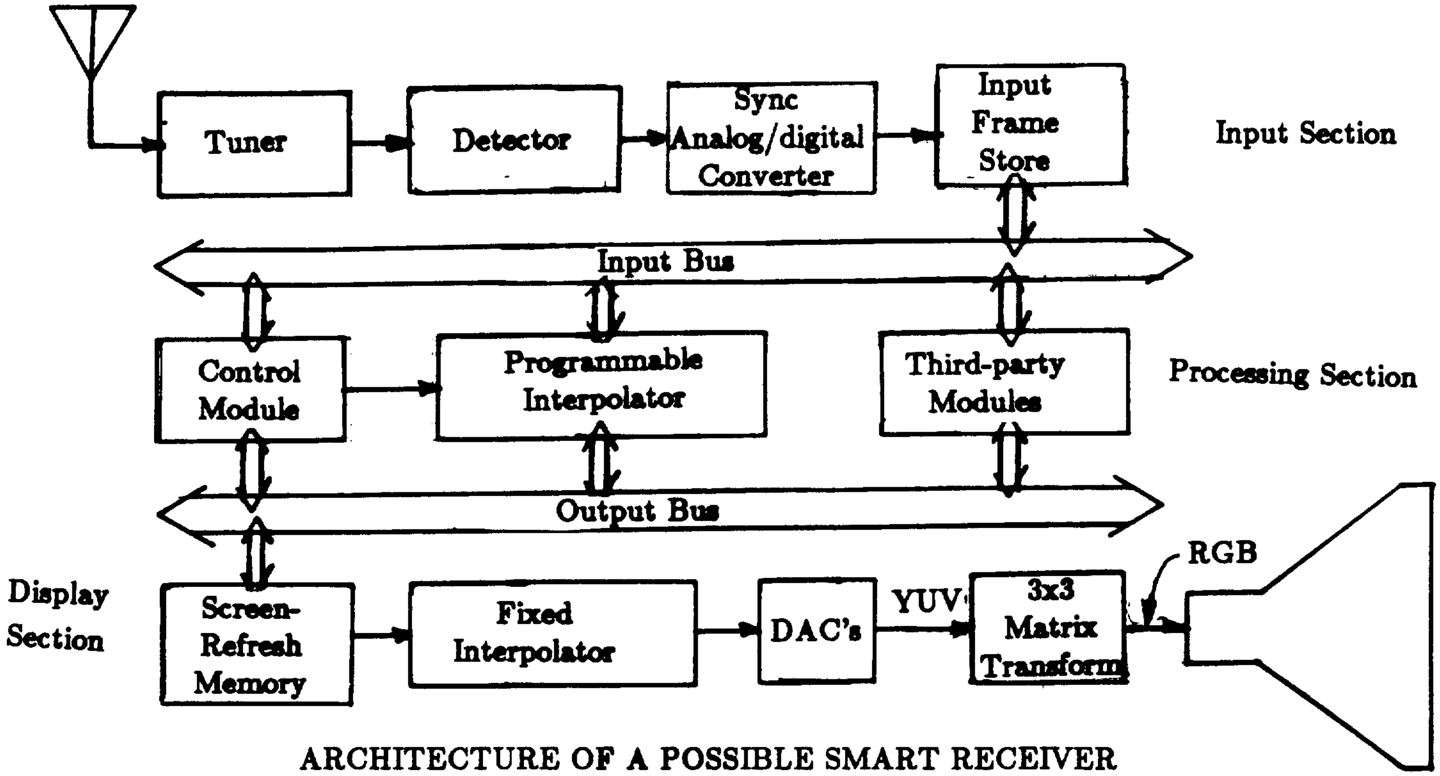

MIT-ATRP HDTV SIMULATION SYSTEM

Sequences of output images are computed and stored on the disks. For NTSC output, the images are written on the Ampex 1" VTR one frame at a time. The tape is then played back at normal speed for viewing. For upconverted output, the images are written on the Fujitsu parallel-transfer disk(s), buffered in the Dataram, upconverted in the 3-d interpolator, and viewed on the 1000- or 2000-line monitor at 60 fps progressively scanned. The 1000-line system has been in operation since March 1987; the 2000-line system will be in operation by January 1988.




NTSC-COMPATIBLE EDTV SYSTEM

Space is made for enhancement signals by using 25% of the height of the NTSC frame, leaving a 16:9 area for the image. Thus, both the NTSC and EDTV images have the same aspect ratio. The enhancement signals are scrambled to the extent practical, and other steps may be taken to make their appearance less disturbing on normal receivers. The NTSC sound carrier is continuous through the enhancement region, but otherwise this time is used in an efficient manner, as described for the noncompatible system, to transmit augmentation signals. These are used to increase the vertical and horizontal resolution, but not the diagonal resolution. Some portion of the channel capacity in the enhancement region may be used for audio, data, and/or progressive scan of a low horizontal frequency portion of the 2-d spectrum.




ARCHITECTURE OF A POSSIBLE SMART RECEIVER

The input and display sections are fixed, while the processing section is programmable under the control of a small amount of digital data transmitted along with the signal. This section can also be upgraded by adding or exchanging modules, some of which could be offered by third parties. A more advanced concept would put the detector in the processing section, in which case its output could be a digitized IF signal extending from perhaps 1 to 7 MHz. This would permit programmable digital detection of signals with multiple carriers, such as in the Philips or NBC/Sarnoff systems. Other configurations are possible for the display unit, which probably should use the mixed highs or luminance/chrominance representation rather than RGB.

*Psychophysics and the Improvement of Television Image Quality*

Society of Motion Picture and Television Engineers Journal (SMPTE)
August 1984

*Improved Television Systems: NTSC and Beyond*

SMPTE Journal
August 1987

William F. Schreiber

These papers are intended for scientists and engineers working on the development of advanced television systems. They should be fully understandable by any recent engineering graduate and understandable in large part by engineers and scientists in any field.

The first paper gives the technical background in visual perception and in signal processing that must be considered in designing improved television systems. It discusses the various limitations to picture quality in existing systems. In a general way, methods of applying these fundamental ideas in TV system design are discussed.

The second paper discusses the specific problems of existing television systems and how they limit picture quality. Methods of improving quality in present systems, both with and without extra channel capacity, are discussed. Requirements for totally new systems are given, and ways to design them are discussed in the light of fundamental principles of psychophysics and signal processing. Examples are given of HDTV within a single 6-MHz channel, HDTV using the digital transmission standard of CCIR Recommendation 601, and super-HDTV (far exceeding typical 35-mm image quality) at about 500 Megabits/sec.

SMPTE Journal

# Psychophysics and the Improvement of Television Image Quality

**By William F. Schreiber**

*Society of Motion Picture & Television Engineers*

**Journal Award**
TELEVISION

*For the most outstanding paper originally published in the Journal of the SMPTE during the year 1984 is awarded to*

[redacted]illiam [redacted].[redacted]chreiber

"PSYCHOPHYSICS AND THE IMPROVEMENT OF TELEVISION IMAGE QUALITY"
AUGUST, 1984, PAGE 717

October 29, 1985

President

Secretary

Reprinted from the *SMPTE Journal*
August 1984 Issue, Volume 93, Number 8

# Psychophysics and the Improvement of Television Image Quality

**By William F. Schreiber**

Worthwhile improvement in television image quality is obtainable by signal processing at the receiver. However, improvement to the level demonstrated by NHK requires a large bandwidth expansion if only straightforward means are used, such as increasing the line and frame rates. This paper discusses a number of methods for obtaining maximum quality for a given bandwidth. Some of these methods take advantage of visual psychophysics, which is reviewed. Others deal with the special characteristics of TV cameras, displays, and scanning patterns. Quite complicated signal processing, expected to become practical in the next few years, is proposed to improve system performance.

Thirty years' experience with the NTSC and PAL TV systems has demonstrated the general soundness of the original concepts and the appropriateness of the chosen parameters. Despite the stringent constraints of compatibility with the then-existing monochrome system, picture quality has proven acceptable, the hardware sufficiently inexpensive and reliable, and a large industry has arisen based on this technology.

## Introduction

### *Motivation*

A number of forces have developed for changes in these systems with a view toward improving picture quality. One is the rapid increase in the variety and capability of semiconductor devices, especially memory, and the accompanying decrease in cost. Much more sophisticated signal processing is thus becoming feasible. Frame memories will probably become practical in receivers before the end of the decade. Many other improvements, such as comb filters and digital demodulation, are already practical. Other possibilities arise from digitization of post-production, which promises greater convenience and flexibility for the producer, more complicated effects, higher signal-to-noise ratio (SNR), and perhaps precorrection for certain degradations likely to be produced by channel and receiver.

The strongest impetus for improvement has undoubtedly come from the demonstration of the Japanese (NHK) HDTV system.[1] While it is not surprising that better pictures can be obtained with four to five times the bandwidth, impressive technological virtuosity was exhibited by the development of the system components, particularly the camera and picture tubes. The sight of vastly improved images, comparable to 35mm theatre quality, on real TV equipment, has whetted everyone's appetite for more improvements, but preferably with less increase of bandwidth.

The path to the practical application of these potential improvements is hardly clear. There are few channels suitable for the NHK system* and there is a serious question as to whether, or by what means, a new system ought to be made compatible. Many possibilities for improvement have been demonstrated which do not require so much bandwidth.[2] Digitization for such a system would be considerably more difficult and expensive than in the case of NTSC.

The principal purpose of this paper is the discussion of methods, based on visual psychophysics and signal processing, by which maximum picture quality can be obtained for whatever channel capacity is provided. It is recognized that there are many other important considerations in the design of new TV systems, such as removing the defects of NTSC, but they are not discussed here.

### *TV as Visual Representation*

In a sense, the TV system substitutes for directly viewing the original scene; hence its success in that role can be used as a measure of its performance. True "presence" is unattainable with any currently proposed system, not only because of the limited spatial and temporal bandwidth and field of view, but because of the two-dimensional (2-D) representation of a three-dimensional (3-D) scene. A truly serious limitation is the use of a single monocular camera of fixed gaze, perhaps panning to track a (single) moving object, while the viewers are many and are constantly moving their eyes over the scene. Finally, although the large-area color reproduction is often excellent, the dynamic range of a cathode-ray tube is far below that of most outdoor and many indoor scenes. The increased definition and field of view of the NHK system are steps in the right direction. However, its motion rendition is bound to be poorer than at present since the field of view is larger and the frame rate is the same.

### *The Potential Contributions of Psychophysics*

Psychophysical principles were applied in the development of both the monochrome and color NTSC proposed standards.[3] The relative horizontal and vertical resolution, the frame rate, the use of interlace, and the overall image quality goal were all selected in this manner in 1941 for the monochrome system. In the color deliberations, the primary psychophysical contributions concerned the representation of the color signal as luminance plus lower-resolution chrominance. Nonvisibility of the color

Presented at the Society's 18th Annual Television Conference in Montreal (paper No. 18-3) on February 10, 1984, by William F. Schreiber, Massachusetts Institute of Technology, Cambridge, MA. This article was received February 10, 1984, and also appears in *Television Image Quality*, published 1984, SMPTE.


* Some examples are direct broadcasting from satellites (DBS), cable TV, and fiber optics.

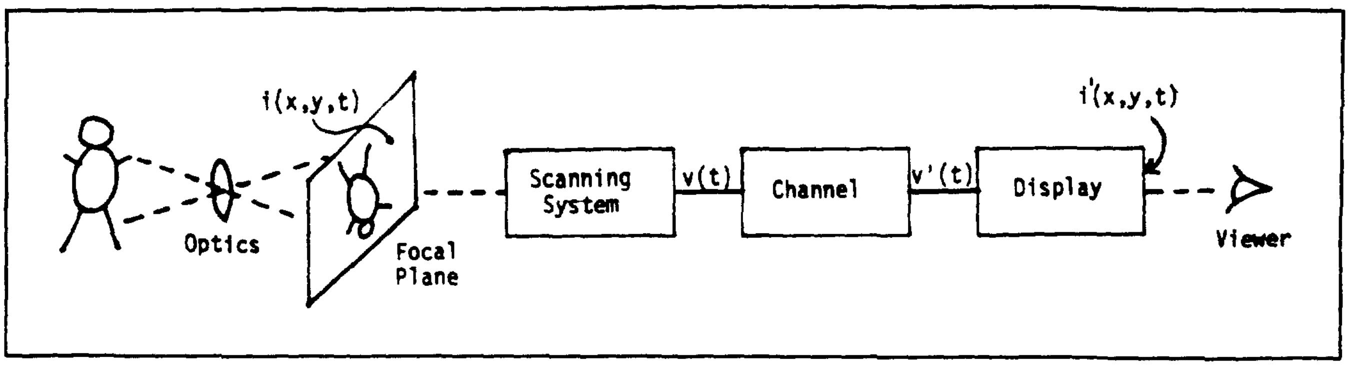



**Figure 1. A generalized TV system.**

subcarrier was more a hope than a fact, and the desired noninterference between chrominance and luminance never did exist, in general. There was no justification in psychophysics for the gross disparity between vertical and horizontal chrominance resolution. It is true that a number of the color-related problems of NTSC were less visible because of the properties of the then-existing transducers. In any event, virtually all contemporary proposals abandon the nonreversible mixing of chrominance and luminance.

One hope for future improvement rests on the considerable body of evidence that neither the NTSC nor the NHK system makes maximum use of the luminance bandwidth. The sampling theorem states that a certain 3-D bandwidth (the Nyquist bandwidth) should be recoverable "exactly," given the vertical and temporal sampling frequencies and the signal bandwidth. Yet, through a combination of factors, the system throughput, at the upper (3-D) limits of the Nyquist bandwidth, is much less than 100%. Furthermore, simply increasing the response would not increase perceived quality in most cases, since certain defects would become more obvious. A number of demonstrations have shown that much better pictures can be produced from the existing transmitted signal simply by up-converting the line and/or frame rate, thereby decreasing flicker and the visibility of the line structure.[4] Wendland has proposed the use of spatially interlaced sampling to accord more closely with the angular dependence of visual acuity,[5] and Glenn has proposed exploiting spatio-temporal interactions for much the same purpose.[6]

In this paper, we shall first describe the television process from the viewpoint of linear signal transmission theory, the input being the collection of illuminated objects before the camera and the output being the picture display as perceived by the viewer. We shall then review some psychophysical data that characterize visual response under controlled (and, unfortunately, rather artificial) conditions. With this background, we shall calculate the required channel capacity for a variety of idealized systems, and show that no straightforward system can give greatly improved quality without unreasonable bandwidth expansion. Finally, we shall discuss a number of alternatives to current TV system design that exploit more thoroughly what is known about human vision. Most of these proposals involve signal processing considerably more complex than now used. Thus they may not become economically feasible for a number of years. When they do become practical, however, they promise a much better quality/bandwidth ratio than is now achievable. We shall not discuss the additional improvement that might be attained by statistical coding.

## The TV Chain as a Linear System

### *A Generalized TV System*

As shown in Fig. 1, light from the scene before the camera is caused to form an image, $i(x,y,t)$, in the focal plane. We call this image the "video function." It is a vector for colored images. The purpose of the system is to produce a modified version, $i'(x,y,t)$, on the display device for viewing.

The video function is converted to a video signal, $v(t)$, by a scanning process operating on the charge image developed by the camera. A simple view of this process is that the signal produced from each point of the focal plane is proportional to the intergrated light power that falls on the point between sampling times. The video signal is further processed by the channel (modulation, filtering, digitization, transmission, etc.), producing a modified signal, $v'(t)$, to be applied to the display device. The display process can be thought of as tracing out, on the viewing surface, a scanning pattern (raster) like that in the camera, in which an amount of energy is emitted at each point of the raster proportional to the light energy collected at the corresponding element of the camera focal plane. In practice, the emitted energy is spread out over a time interval, almost always much shorter than one frame time.

This description reveals a significant difference between the original and reproduced video functions. The former is continuous in space and time, while the latter is highly discontinuous. If the output were continuous, the system could be characterized simply by its spatio-temporal frequency response which could then be compared with the corresponding sensitivity of the human visual system (HVS). This space-time discontinuity is the cause of much of the inefficiency in utilization of the channel capacity. Simple-minded elimination of the sampling structure by blurring, the use of long-persistence phosphors, or by viewing from a distance, attenuates important components of the transmitted signal as well as the structure.

### *The Special Problem of Interlace*

Since 30 frames/sec, progressively scanned, produces totally unacceptable large-area flicker, interlace was introduced early in the history of TV development, doubling the flicker rate to 60 Hz while preserving the full number of lines in the frame. The only condition under which 30-Hz large-area flicker can result with interlace is when the average brightness of odd and even fields is significantly unequal, a rare event.

Interlace has problems as well as

advantages. It was recognized at an early date that with phosphors of persistence short enough not to cause interframe blurring, vertical motion could often produce a display with half the number of scan lines. Horizontal motion ought to produce serrated vertical edges, but usually does not because of camera integration. It was not generally recognized that for viewing distances at which the lines can be clearly resolved, the interline flicker rate is 30 Hz, easily seen as a shimmer. A side-by-side comparison of interlaced and noninterlaced images (the latter requires twice the bandwidth, of course) makes these differences very obvious.

Even at viewing distances at which the line structure cannot be resolved, 30-Hz flicker is clearly visible in interlaced pictures in areas having significant vertical detail. Flicker occurs when odd and even lines are sufficiently different at *any* resolvable spatial frequency. This flicker can be eliminated either by reducing the vertical resolution of the camera and/or the display, or by integrating over a full frame by some temporal averaging device. With any such method, spatio-temporal resolution is reduced.

In the light of these considerations, the subjective effect of interlace has never been fully investigated, since vertical resolution, whose role has only been appreciated recently, was not adequately controlled. Even so, Brown's early study concluded that for a 225-line TV image at 50 fL, viewed at eight times picture height, interlace produced a subjective increase in vertical resolution of only 24% (36% at 40 fL, and a mere 6% at 100 fL) in the line number compared with a progressively scanned image at 60 Hz.[7] A similar NHK study in 1982 showed an increase of 20% for a 1500-line picture viewed at two times picture height.[8] These numbers are so much lower than 100%, which would be obtained if interlace "worked," one wonders why it has been thought to be so effective. For present-day scanning standards, interlace clearly produces artifacts which become more troublesome as the vertical resolution is increased and as the image is more closely viewed, while the vertical resolution is increased only lightly.

### *Special Properties of the Camera*

In most camera tubes, the target, which integrates the incident light at each point between successive visits of the scanning beam, is almost completely discharged each field. The integration area, in the vertical direction, thus comprises at least two of the 525 nominal scan lines. A vertical pattern of 262.5 cycles per picture height (cph) is rendered with zero response, and a frequency of even half that is substantially attenuated, to a degree that depends on its phase. Yet the sampling theorem tells us that we ought to be able to use the full bandwidth of 262.5 cph. If, however, the vertical response of the camera is increased, as it readily can be, for example, in laser scanners, we see disturbing interline flicker.

In some modern CCD cameras which have one row of detectors for each scan line, the pairing of two lines of data for each output line is deliberate.[9] In cathode-ray camera tubes, the process is more complex due to the physics of target discharge and the shape of the electron beam.[10] In this case, dark areas are completely discharged by the leading edge of the beam, while bright areas are not fully discharged until passed over by the trailing edge. The resulting geometrical distortion and small-area tone-scale distortion are not very serious. More important is the fact that, as ordinarily operated, the vertical resolution of camera tubes is much less than the horizontal resolution (expressed as lines/mm on the target). This is fortuitous, since higher vertical resolution would make interlace even less acceptable. However, the result is that by employing interlace, we have sacrificed a significant portion of the theoretically available vertical definition, and with it, much of the expected benefit.

### *The Picture Tube*

Cathode-ray display tubes are essentially linear; thus the integrated light output at each point can be found by convolving an ideal (zero spot diameter) raster with the beam cross section, which is generally Gaussian. With such a shape, the elimination of line structure by defocussing (or by blurring in the eye) also blurs the image. In color tubes, an additional factor is the structure of small phosphor spots. In a 19-in. diagonal shadow-mask tube with 0.31-mm triad spacing, only about four triads are available for each picture element in the NTSC system. This is bound to introduce a great deal of spatial high-frequency noise, to which, fortunately, we are not very sensitive. However, the channel SNR for AM transmission is uniform with bandwidth and therefore does not take advantage of this phenomenon.

### *The Channel*

In present-day systems, the purpose of the channel is to reproduce at the picture tube the output of the camera tube with perhaps some minor amount of processing. Of course, noise is invariably added in the process and there may be some loss of bandwidth. As pointed out later, the most probable source of major improvement would be the introduction of substantial signal processing between camera and channel and between channel and display. Since the second processor must be cheap, that is the location of the significant technological challenge.

## The Psychophysical Background

### *Normal Seeing*

The HVS, presumably as a result of evolution, is well adapted to rapidly deriving a large amount of useful information from the scene before the observer. This scene, 3-D, variously illuminated, and moving, produces slightly different 2-D images on the retinas of the two eyes. The eyes are in constant voluntary motion over the scene, both by head motion and by rotation in their sockets. They also execute small involuntary motions which have been found to have an important, even essential, role in vision.[11]

The retina consists of a matrix of receptors of two kinds — cone cells which exclusively cover the central 2° (the fovea) and whose density decreases away from the axis, and rod cells whose density is maximum 15° from the center. Cones are responsible for the high visual acuity on axis and for color vision at normal (photopic) levels. The rod cells, which are much more sensitive, provide off-axis low-light-level (scotopic) sensitivity but have much lower spatial resolution.

The sensitivity of each cell depends on its state of adaption and on the excitation of its neighbors. The sensitivity is characterized by both a static (input-output) function and a frequency response. Spatial resolution of point objects is roughly equal to cell spacing, but because of cooperation of retinal receptors, resolution of long parallel lines is much finer than the cell spacing. The discrete nature of the retinal mosaic is never obvious in normal vision. Much visual processing is carried out on the retina itself, but additional processing occurs at higher levels of the nervous system.[12]

### *Characterization of Visual Response*

Learning about vision is a frustrating study, since the vast literature would take years to master, yet data are lacking on many points that are vital to the design of efficient systems. Although all aspects of the visual sense are remarkably interdependent, it is customary to begin by discussing its performance along separate axes.

#### *Contrast Sensitivity*

By contrast sensitivity, we mean the visual response as a function of luminance, although what is usually measured is the just-noticeable difference between near-equal luminances displayed side by side or one after the other. Obviously, temporal or spatial separation is essential for measuring contrast thresholds, so that it is quite impossible to separate contrast sensitivity completely from these other variables.

With the usual test field (Fig. 2), the observer is allowed to adapt to the surround, $L$, and then the smallest discernible $\Delta L$ is found.[13] The result of this measurement (Fig. 3) shows that $\Delta L/L$ is nearly constant over five decades. We can thus see over an enormous luminance range, given time to adapt. The constancy of $\Delta L/L$ is called the Weber-Fechner law, the fraction being as small as 1% under optimum viewing conditions.

In the more normal situation — observing actual scenes or their reproductions — the degree of adaptation is much less. If we now measure $\Delta L/L$ as a function of the adapting luminance $L_0$, using a target such as that of Fig. 4, we find that the operating dynamic range is much smaller. More significant is the appearance of the central patches as a function of the

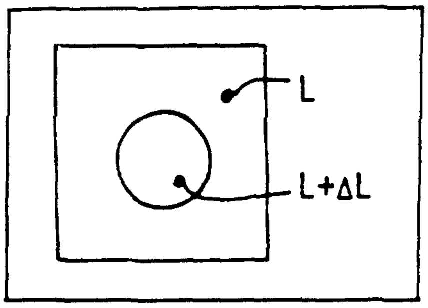



**Figure 2. Contrast sensitivity target.**

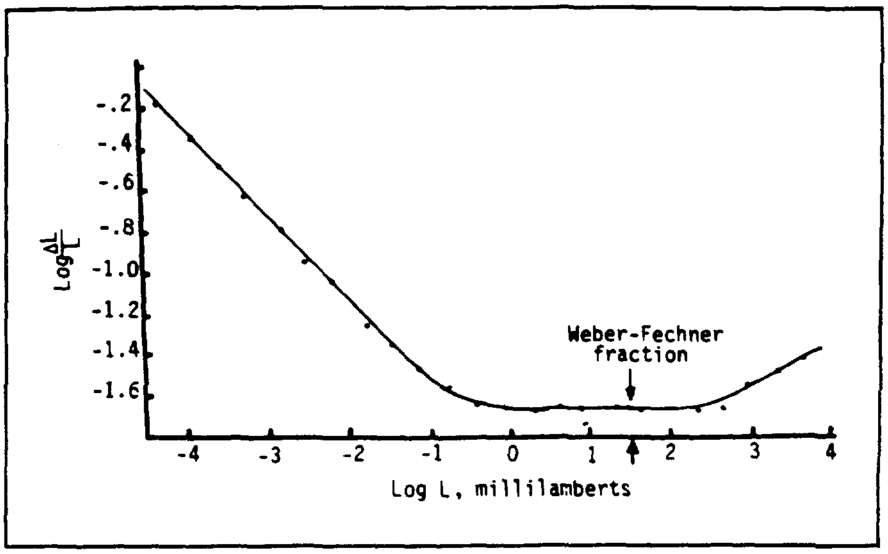



**Figure 3. Constrast sensitivity data of Koenig and Brodhun, 1884. (Quoted by Hecht, *J. Gen. Physiol.* 7:421, 1924.)**

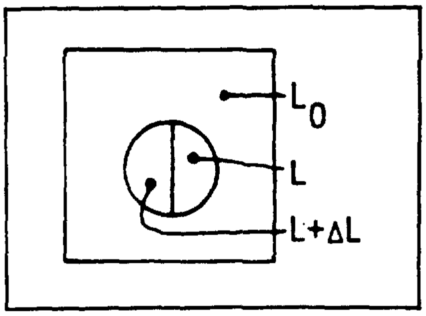



**Figure 4. A more realistic contrast sensitivity target.**

relative brightness. When the surround is about 100 times brighter than the central area, the latter looks black, no matter what its actual luminance, while in the reverse case it looks white.[14] When the central area is sufficiently intense, it appears to be a light source rather than an illuminated surface.

For picture reproduction, this means that nearly four decades of dynamic range are required to give the visual impression of a real high-contrast scene, such as outdoors on a clear day. This condition is approximated by optical projection from film with good equipment in a perfectly dark room. Under all other conditions, such as TV displays, the dynamic range must be compressed. Although this can be done so as to give pleasing results in terms of brightness and contrast as those terms are normally used, it is very hard to impart realism.

#### *Temporal Frequency Response*

Flicker and motion rendition are associated with the temporal response factor, so it is of great importance and has had the attention of psychologists for many years. The "purest" method of measurement (least contaminated by other factors) is to superimpose a sinusoidally fluctuating component on a constant luminance and to use a very wide field with defocussed edges. The definitive measurement has been made by Kelly.[15] The most interesting aspect of his results is that over a significant range of temporal frequencies, the HVS is a differentiator (Fig. 5), not an integrator. Flicker in this range is very noticeable. It is quite evident that at 25 or 30 Hz, flicker is almost always present. At 50 or 60 Hz, it is present in very bright images. To avoid flicker in the worst case, which is at the edges of a bright, wide-field display, 80 or 90 Hz might be needed. Note that peripheral flicker is sometimes seen in wide-screen motion pictures, where the flicker rate is usually 72/sec.

#### *Spatial Frequency Response*

Visual acuity — the ability to see sharply and resolve small details — is one of the most obvious aspects of vision. Although threshold contrast is often measured as a function of spatial frequency using square-wave gratings at various angles, the results are easier to interpret if sine-wave gratings are used.[16] A variety of indirect methods have also been used, in which transient response[17] or response to filtered random noise[18] has been measured. Despite the hazards of applying linear analysis to such a nonlinear system, all the results are similar to those shown in Fig. 6. Remarkably, the spatial characteristic also shows a differentiation region, one of the effects of

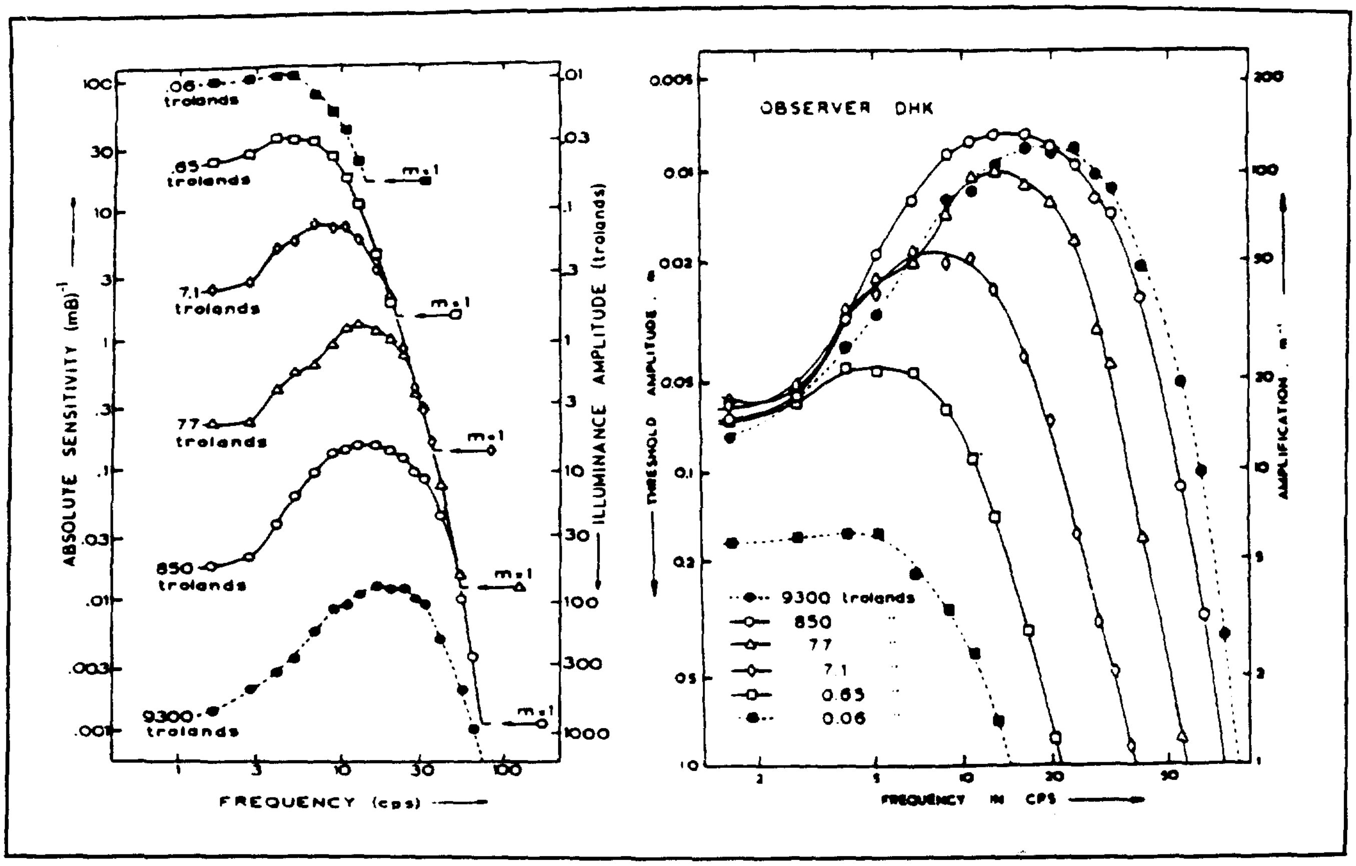



Figure 5. Kelly's temporal data, plotted two ways. (From D. H. Kelly, "Visual Response to Time Dependent Stimuli. I. Amplitude Sensitivity Measurements," *J. Opt. Soc. Am.*, Vol. 51, No. 4, 1961, pp. 422-429.)

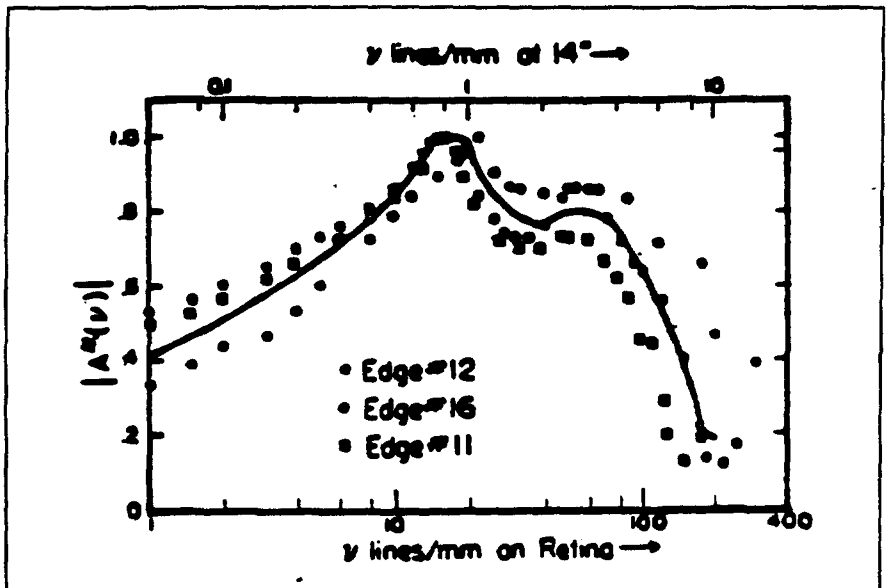



Figure 6. Spatial frequency data of Ref. 17. (E. M. Lowry and J. J. DePalma, "Sine Wave Response of the Visual System," *J. Opt. Soc. Am.*, Vol. 51, No. 10, 1961, p. 474.)

which is to sharpen images significantly. It is thought that this effect is due principally to neural interaction on the retina. Note that there is some response up to 30 or more cycles per degree (cpd). For a 90° display, 2700 cycles, or 5400 picture elements, would be required for absolute invisibility of the scanning structure. For the lines in an NTSC picture to disappear completely, only a 16° field can be covered.

In the vertical and horizontal directions, the spatial frequency response is almost equal, but at 45° it decreases by a factor of 2 or more. This is the main reason why half-tone patterns are usually at 45°. It is also the basis of proposals for interleaved sampling.[19] Whether this would be advantageous is hard to say. Baldwin carried out a very careful experiment to determine the effect on picture quality of varying the relative horizontal and vertical resolution.[20] For pictures of 56-in.$^2$ area with about 36,000 resolvable elements viewed at 30 in., *the just-noticeable degree of asymmetry was 2.5:1*, despite the equal horizontal and vertical limiting resolution of the eye. Thus it is not obvious that interleaved sampling would improve image quality, although it might make structure less visible.

*Spatio-Temporal Interactions*

Measuring the combined effect of spatial and temporal fluctuations is more difficult, and a wider variety of methods can be used. There is reasonable agreement that peak sensitivity is at about 2 Hz and 2 cpd, with integration at higher frequencies and differentiation at lower frequencies. This has been modelled as the difference between an excitatory and an inhibitory response, an interesting point, but not of direct value to the system designer.[21] There is evidence that the shape is somewhat more complicated, but the essential result is that the derived passband in 3-D frequency space is not cubical, but more nearly ellipsoidal. No one seems to have repeated Baldwin's experiment for spatio-temporal resolution. Just because the limiting spatial and temporal frequencies have been shown to be inversely related, does not prove, in a system where the signal components are well below the limiting frequencies, that the system bandwidths ought to be so related.

*Masking*

Of great interest to the psychophysicist, and in this case of equal interest and value to the system designer, is the phenomenon of masking. In all sense modalities, response to particular kinds of stimuli is reduced significantly by the presence, in the immediate spatio-temporal neighborhood, of similar stimuli. In the case of large, uniform, slowly changing scenes, we call this phenomenon adaptation. It is of great value because it enables us to see well under a wide variety of conditions.

A similar phenomenon occurs for stimuli of similar spatio-temporal content. Exposure to a spatial grating reduces sensitivity to gratings of similar spatial frequency seen just afterwards or even just before.[22] Exposure to a temporal sinusoid reduces sensitivity to sinusoidal flicker of like frequency.[23] The presence of "activity" (sharp edges or fine detail) reduces noise sensitivity in nearby areas.[24] An example of the latter in the space domain is the much lower visibility of additive random noise in detailed or "busy" image areas and its much higher visibility in relatively blank areas. For this reason SNR, even weighted according to the variation of noise visibility with frequency, is a very poor indicator of image quality. Simple images require a much higher SNR than complicated ones for the same visual quality.

A related phenomenon is the masking of detail in a new scene by the presence of a previous scene. Repeating an earlier experiment by Seyler[25] in our own laboratory, we found that a new scene could be radically defocussed and then refocussed with a time constant of 0.5 sec, without visible effect. Recently, Glenn has demonstrated that it takes about 0.2 sec to perceive higher spatial frequencies in newly revealed areas.[26] This effect is nature's gift to temporal differential transmission systems, since it allows new scenes to be built up over a period much longer than a frame.

*Motion Rendition*

Americans believe that it was Edison† who discovered that the illusion (sic) of motion could be produced by viewing a rapid sequence of slightly different images. This is the "phi motion" of psychology, in which the successive flashing of two small lights, with the appropriate time and space separation, makes it appear that a light moves from the first position to the second.[27] Should the angular jump be too large or the interval too long, the motion effect is discontinuous, and in some cases, can even be retrograde. We have all seen wheels standing still or even moving backward. This stroboscopic effect, which has its uses, of course, is an example of temporal aliasing. Like other kinds of aliasing, it is but one possible defect that should be traded off against others for optimum image quality. The smoothness of motion is directly related to filling the gaps between successive positions. The degree of temporal bandlimiting required to preclude temporal aliasing absolutely, has the effect of blurring moving objects.[28] Especially in low-frame-rate systems, it may be preferable to show a sequence of sharp still images rather than a continuously moving image so blurred as to be useless.

Careful observation shows that motion is generally smoother in TV than in motion pictures. This is because the TV system actually takes 60 pictures/sec, as compared to 24 for film. In addition, most TV cameras integrate for the full $1/60$ sec, while all motion-picture cameras use exposure times of less than (and sometimes very much less than) $1/24$ sec.

Objects that move across the retina while the eyes are fixated elsewhere are blurred by the temporal upper frequency limit of the HVS. The same thing happens in TV cameras, which is harmless unless the observer happens to be tracking the object. In that case the TV (or motion-picture) representation is disappointing. There is thus no way the TV camera can satisfy the entire audience when the scene contains two or more important moving objects.

† No doubt other countries have their own favorite inventors of motion pictures.

*Color*

Color is not a principal preoccupation of this paper since colorimetry is quite satisfactory in existing systems. In a new system design not constrained by the requirement of compatibility, however, there are several simple ways of adding color to a monochrome signal. Based on the lower required spatial color resolution, these methods increase the channel capacity by 20% or less.[29] More complicated systems decrease the color penalty even more.[30] At these incremental levels, the color picture, with slightly lower luminance resolution, is usually far superior in perceived quality, almost however measured, to the monochrome picture with slightly higher luminance resolution.[31] Thus, the addition of color can be viewed as a valuable way to *decrease* the total channel capacity for a given subjective quality.

It is also possible, but not yet demonstrated as far as we know, that the required temporal bandwidth for color is less than for luminance, in which case an additional possibility for compression would be available.[32]

## Performance Goals for TV Systems

*Perfection*

A perfect system can be defined as one that gives a convincing illusion of reality. This probably would not require 3-D reconstruction, as might be done holographically. A very wide field of view is quite effective.[33] Assume that 90° vertically and 180° horizontally would be enough. The question, then, is the required resolution. A frame rate of 100/sec would certainly prevent flicker, but even that would not keep rapidly moving objects in focus. Using 50 cpd as the upper perceptual limit, a raster of about 9000×18,000, or 162 million samples/frame would suffice, for a total rate of 16 billion samples/sec. Of course such a signal would have very high redundancy and could be greatly compressed. Nevertheless, the obstacles to constructing such a system are insurmountable at present.

*Idealism*

An ideal system, for our purpose, can be based on resolution parameters so high that raising them would not materially improve quality. We would, however, accept a more limited field of view and the motion rendition obtainable at 60 frames/sec. For a 45° × 90° field of view and a sampling density of

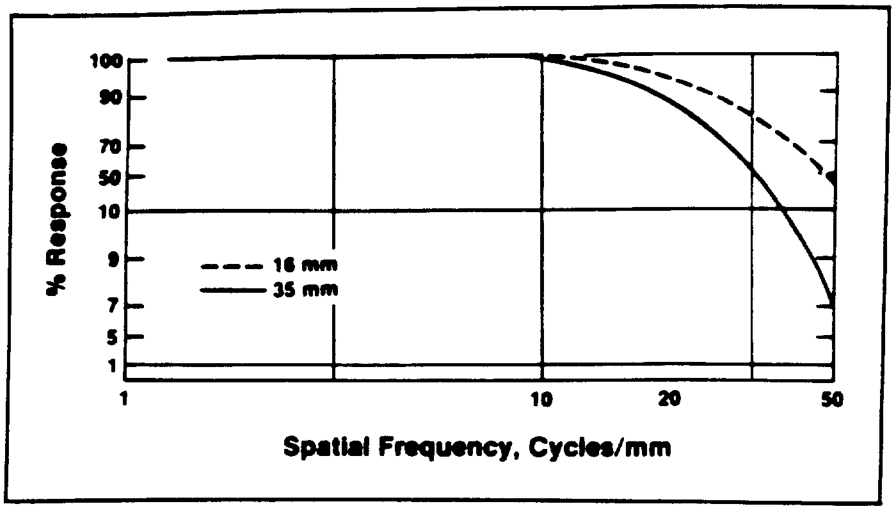



Figure 7. Overall response of film system. (From R. C. Sehlin, et al, *SMPTE Journal*, December, 1983.)

12/mm at normal viewing distance (30 cm), which is considered excellent quality for continuous-tone color prints, we would have a more modest 3000×6000 raster — a mere 18 million samples/frame, or 1 billion samples/sec. To get the full benefit from this resolution, we should probably raise the frame rate somewhat — perhaps to 80/sec, for a rate of 1.25 billion samples/sec. These pictures would give the effect of looking at the real world through a large window, except that if we were to track rapidly moving objects (those that move across the screen in less than 4 sec or so), we would see a definite loss of resolution.

### *Theatre Quality*

The term "theatre quality" is very poorly defined, since film is getting better and better,[34] and we now know how to make nearly diffraction-limited optics. For the sake of discussion, for 35mm film with a frame height of 18mm and 30 to 50 line pairs/mm assumed for the effective resolution limit, 1080 to 1800 lines and 1.5 to 4.3 million samples/frame would be required. This is in accordance with the NHK experience.

Equating TV and film quality is not simple and certainly requires careful subjective testing. The spatial frequency response of film and optical systems tends to fall monotonically, starting at a low spatial frequency (Fig. 7). Television systems have a rather well-defined upper frequency limit, but within the passband we are free to use almost any characteristics we wish. A considerable degree of sharpening is possible and is routinely used in electronic-based graphic arts systems.[35]

## Vision-Based Design

### *Spatial Filtering, Sampling, and Interpolation*

Input and output still images are inherently spatially continuous. When represented by an array of numbers, the continuous-discrete and discrete-continuous conversions can have a significant effect on image quality. Since analog TV is sampled only in the vertical direction, this section applies principally to processing designed to give maximum vertical sharpness without artifacts.

A basic problem with discrete imaging systems lies with the sampling theorem, which states that the recoverable signal bandwidth (Nyquist bandwidth) is one-half the sampling frequency. At the transmitter, the bandwidth should therefore be limited to the Nyquist value to prevent aliasing, but there is no obvious mechanism for accomplishing this in a TV camera. Furthermore, if we somehow did implement such a filter, the ringing associated with sharp horizontal edges would be unacceptable. There *is* an optimum filter and its implementation will be discussed shortly.

At the display, it clearly would be desirable to eliminate the scan lines. They are obtrusive and, due to the masking effect, suppress the high-frequency structure to some extent. Achieving this result by defocussing the more or less Gaussian scanning beam causes noticeable loss of sharpness. Relying on the filter of the HVS produces a similar effect, although with less loss of sharpness. In any event, the effect of the HVS is strongly dependent on the angular subtense of the scan lines at the eye (Fig. 6). When viewing an NTSC picture at 4H, the line structure is 34 cpd. At these spatial frequencies, visual response drops about 18 dB/octave. Thus the line structure is attenuated 18 dB compared to the signal components at the upper end of the Nyquist band. However, to achieve this separation, the signal components are also attenuated substantially. Viewed at 2H, the relative attenuation is only about 12 dB.

Vertical filtering can be done effectively by operating both camera and display at a substantially higher line rate, and interposing processing elements between camera and channel and channel and display (Fig. 8). At both camera and display, this could give enough vertical samples to implement the appropriate digital filter. At the display, such up-conversion would also raise the line rate to a point where the HVS could more easily separate the structure from the image. Incidentally, but perhaps importantly, high-line-rate operation of the camera might well ameliorate the problems discussed above due to the nonlinear target discharge, especially with progressive scanning. The application of similar methods to still pictures has resulted in a channel capacity saving of as much as 40% for the same perceived quality, as compared with simple-minded methods.[36]

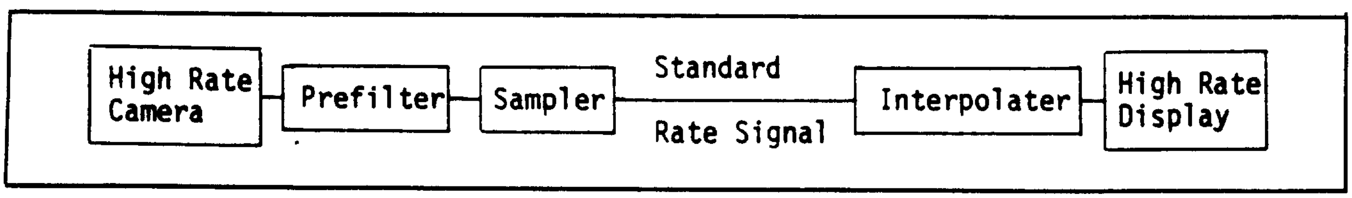



Figure 8. The modified TV chain.

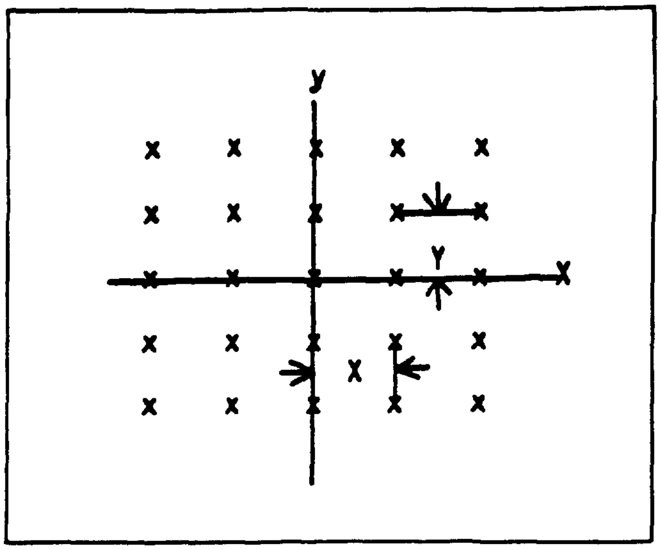



Figure 9. Cartesian sampling.

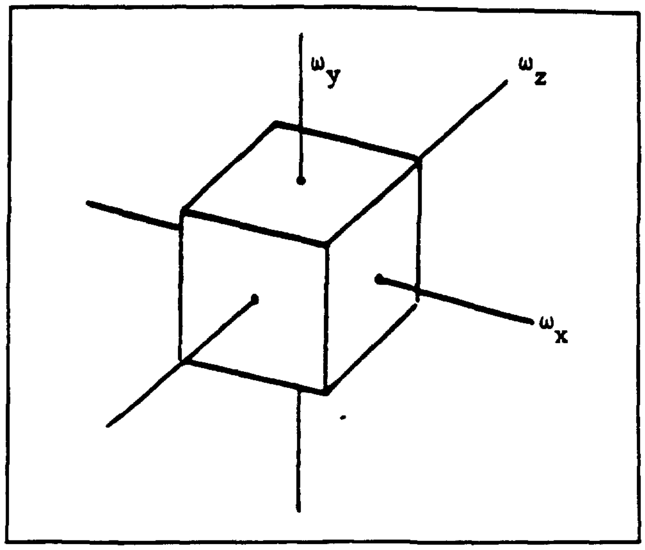



Figure 10. Cartesian spectrum.

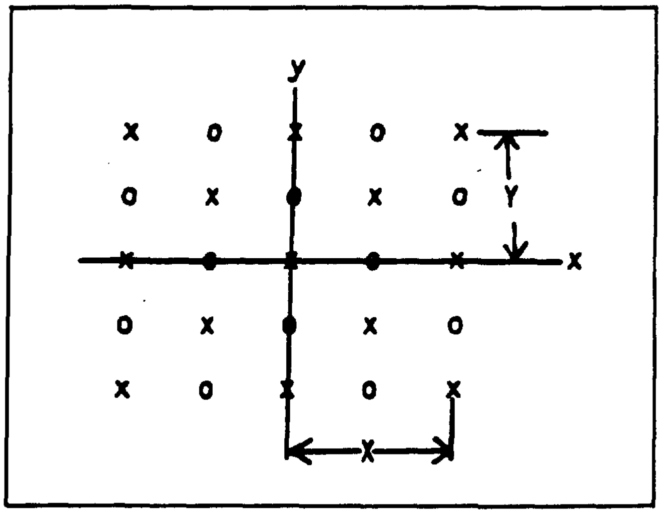



Figure 11. Temporal interleaving: *x*=even fields; *o*=odd fields.

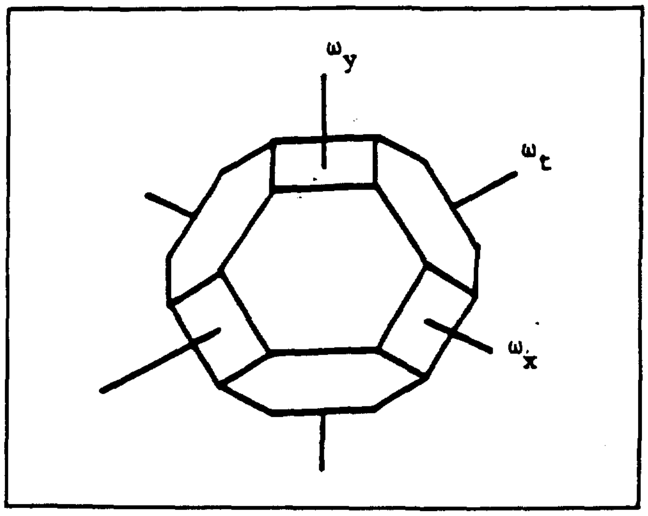



Figure 12. Alias-free bandwidth for temporal/spatial interleaving.

### *Temporal Filtering, Sampling, and Interpolation*

This situation is analogous to that caused by spatial sampling. In this case, the sampling theorem tells us that 30 frames/sec results in a 15-Hz Nyquist bandwidth. To avoid aliasing (jerky or stroboscopic motion), we should low-pass filter before sampling. Since a camera that integrates perfectly for the frame (or field) time is hardly an ideal filter, we could tailor presampling filters much more accurately if the camera operated at four or five times the frame rate. At the display, the extra samples would have a similar effect, but in addition, as in the spatial case, would raise the rate of the flicker so that it could more easily be separated from the baseband by the HVS.

### *Three-Dimensional Processing*

The spatial and temporal processing discussed above could be combined so that the filters of Fig. 8 would be 3-D. Such filters require frame and line stores and can only be implemented digitally. In such a TV system, discrete in all three dimensions, the question of the sampling pattern of the channel signal,[‡] as well as the corresponding 3-D Nyquist bandwidth, must be dealt with.

The Cartesian pattern of Fig. 9 gives the Cartesian spectrum of Fig. 10, while the interleaved pattern of Fig. 11 gives the odd-looking spectrum of Fig. 12. (Other 3-D patterns are possible.) This pattern trades off spatial and temporal bandwidth in a manner that probably is better than the Cartesian pattern, although observer tests are necessary to be sure. It has higher spatial response at low temporal frequencies and vice versa, and higher vertical and horizontal resolution than diagonal.

Interleaved sampling bears some relationship to present-day interlace, which is used primarily to double the flicker rate. However, we can think of the vertically offset sampling of standard interlace as a means of raising the (time-averaged) vertical resolution for a given vertical scan rate. In this endeavor it mostly fails, for the reasons cited. Interleaved sampling as described here, however, when used in conjunction with appropriate 3 presampling and interpolation filters, has none of the defects of ordinary interlace. It aims for and gets no "free" expansion of bandwidth. In fact, the volume of the 3-D Nyquist bandwidth is identical for all sampling patterns that have the same number of samples per unit $(x,y,t)$ volume. What interleaved sampling does do is to change

[‡] The sampling patterns of the camera tube and display are of little importance since they will not be detected by the viewers.

the shape of the Nyquist bandwidth from Cartesian to one that may be better.

It would be difficult to implement a filter with the response shown in Fig. 12. However, an ellipsoidal impulse response would approximate it and, if Gaussian, would be separable and therefore practical. Since a form of Gaussian filter was found optimum in the studies cited,[20] it is quite likely to work well in this case.

### *Multi-Channel Systems*

There is some evidence that the HVS treats various spatial frequency components of the visual stimulus so differently that an advantage can be gained by separating the signal into two or more channels and using different transmission parameters for each component. This is quite in accord with widely held theories that the visual system is organized in this manner.[37] A number of systems quantize low and high spatial frequencies differently, using a rather coarse quantization in the highs channel.[38] The quantization noise, preferably randomized,[39] tends to be masked by the high-frequency detail.

Glenn has suggested that the high frequencies can be transmitted at a lower frame rate.[6] Although some trade-off between spatial and temporal response is possible by offset sampling, Glenn goes much further, transmitting the highs at only 5 frames/sec. In the case of newly uncovered stationary detail (a new scene, or newly revealed background that emerges from behind a moving foreground object), this seems to work rather well. In the case of detailed objects moving in the scene, the blurring must be much worse than at 30 frames/sec. It is possible that visual acuity in the tracking mode is sufficiently low that this is permissible. Clearly more work needs to be done on this technique, since if successful it would permit substantial saving.

## Conclusion

We have described TV transmission as a problem in the analysis of linear systems. A review of the literature on visual psychophysics as it applies to this formulation has revealed a number of possibilities for the improvement of picture quality in relation to channel bandwidth. These involve 3-D processing at both transmitter and receiver, and, in the latter case, would be practical only with a high degree of circuit integration.

Specific visual problems due to the characteristics of camera and display devices and to the use of interlace have been pointed out. The amelioration of such effects by operating these devices at very high line and frame rates requires rather complicated signal processing, but presents the prospect of significant improvement in the utilization of transmission channel capacity.

## Acknowledgments

The literature in this field is large and growing rapidly, so the list of references should be considered representative and not exhaustive. Of the ideas presented, many have grown out of discussions with colleagues, students, and friends. I have had the assistance of G. Saussy in collecting psychological references. S. Sabri motivated me to write this paper.

## References

1. T. Fujio, et al., "HDTV," NHK Tech. Monograph No. 32, June 1982.
2. K. Lucas, et al., "Direct TV Broadcasts by Satellite," IBA Report No. 116, 1981.
3. D. G. Fink, "TV Standards and Practice" (1941 NTSC), McGraw Hill, 1943; Proc. IRE, Jan. 1954 entire issue (1953 NTSC).
4. C. P. Sandbank and M. E. B. Moffatt, "HDTV and Compatibility with Existing Standards," *SMPTE J.*, May 1983, pp. 552-561.
5. B. Wendland, "Extended Definition TV with High Picture Quality," *SMPTE J.*, October 1983, pp. 1028-1035.
6. W. E. Glenn, et al., "Compatible Transmission of HDTV Using Bandwidth Reduction," IGC paper, unpublished.
7. E. F. Brown, "Low-Resolution TV: Subjective Comparison of Interlaced and Non-Interlaced Pictures," *Bell Sys. Tech. J.*, Vol. 46, January 1967, pp. 199-232.
8. T. Mitsukashi, "Scanning Specifications and Picture Quality," NHK Tech. Monograph No. 32, June 1982.
9. N. Murata, et al., "Development of a 3-MOS Color Camera," *SMPTE J.*, December 1983, pp. 1270-1273.
10. T. G. Schut, "Resolution Measurements in Camera Tubes," *SMPTE J.*, December 1983, pp. 1287-1293.
11. L. A. Riggs, et al., "The Disappearance of Steadily Fixated Test Objects," *J. Opt. Soc. Am.*, Vol. 43, 1953, pp. 495-501.
12. T. N. Cornsweet, *Visual Perception*, Academic Press, 1970.
13. Data of Koenig and Brodhun (1884) quoted by S. Hecht, *J. General Physiology*, Vol. 7, 1924, p. 421 ff.
14. E. G. Heinemann, "Simultaneous Brightness Induction," *J. Experimental Psychology*, Vol. 50, 1955, pp. 89-96.
15. D. H. Kelly, "Visual Reponses to Time Dependent Stimuli. I. Amplitude Sensitivity Measurements," *J. Opt. Soc. Am.*, Vol. 51, No. 4, 1961, pp. 422-429.
16. F. W. Campbell and J. G. Robson, "Application of Fourier Analysis to the Visibility of Gratings," *J. Physiology*, No. 187, 1968.
17. E. M. Lowry and J. J. DePalma, "Sine Wave Response of the Visual System," *J. Opt. Soc. Am.*, Vol. 51, No. 10, 1961, p. 474.
18. O. R. Mitchell, Ph.D. Thesis, MIT EECS Dept., 1972.
19. B. Wendland, "On Picture Scanning for Future HDTV Systems," IBC, 1982.
20. M. W. Baldwin, "The Subjective Sharpness of Simulated TV Pictures," Proc. IRE, October 1940, pp. 458-468.
21. C. A. Burbeck and D. H. Kelly, "Spatio-Temporal Characteristics of Visual Systems," *J. Opt. Soc. Am.*, Vol. 70, No. 9, 1980; and F. L. Van Nes, et al., "Spatio-temporal Modulation Transfer in the Human Eye," *J. Opt. Soc. Am.*, Vol. 57, 1967, pp. 1082-1088.
22. G. Sperling, "Temporal and Spatial Visual Masking," *J. Opt. Soc. Am.*, Vol. 55, 1965, pp. 541-559.
23. O. Braddick, F. W. Campbell, and J. Atkinson, "Channels in Vision: Basic Aspects," in *Handbook of Sensory Physiology*, Vol. VIII, Springer-Verlag, New York, 1978.
24. A. N. Netravali and B. Prasada, "Adaptive Quantization of Picture Signals Using Spatial Masking," Proc. IEEE, Vol. 65, No. 4, April 1977, pp. 536-548.
25. A. J. Seyler, et al., "Detail Perception after Scene Changes in TV," IEEE Trans. Information Theory, Vol. IT-11, January 1965, pp. 31-43.
26. W. E. Glenn, "Compatible Transmission of HDTV Using Bandwidth Reduction," National Association of Broadcasters, Las Vegas, April 12, 1983, videotape demonstration.
27. R. A. Kinchla, et al., "A Theory of Visual Movement Perception," *Psychological Review*, Vol. 76, 1969, pp. 537-558.
28. J. Korein, et al., "Temporal Anti-Aliasing in Computer Generated Animation," *Computer Graphics*, Vol. 17, No. 3, July 1983, pp. 377-388.
29. W. F. Schreiber and R. R. Buckley, "A Two-Channel Picture Coding System: II — Adaptive Companding and Color Coding," IEEE Trans. on Communications, Vol. COM-29, No. 12, December 1981, pp. 1849-1858.
30. R. R. Buckley, Ph.D. Thesis, MIT EECS Dept., 1982.
31. U. Gronemann, Ph.D. Thesis, MIT EE Dept., 1964.
32. H. de Lange, "Research into the Dynamic Nature of the Human Fovea-Cortex System, *J. Opt. Soc. Am.*, Vol. 48, 1958, pp. 777-784.
33. W. C. Shaw, et al., "IMAX and OMNIMAX Theatre Design," *SMPTE J.*, March 1983, pp. 284-290.
34. M. A. Kriss, et al., "Photographic Imaging Technology for HDTV," *SMPTE J.*, August 1983, pp. 804-818.
35. W. F. Schreiber, "Wirephoto Quality Improvement by Unsharp Masking," *Pattern Recognition*, Vol. 2, March 1970, pp. 117-121. See also W. F. Schreiber and D. E. Troxel, U.S. Patent 4,268,861, May 1981.
36. J. R. Ratzel, Sc.D. Thesis, MIT EECS Dept., 1983. See also W. F. Schreiber and D. E. Troxel, "Transformation between Continuous and Discrete Representations of Images: A Perceptual Approach," MIT Report, submitted for publication.
37. L. Maffei, "Spatial Frequency Channels: Neural Mechanisms," in *Handbook of Sensory Physiology*, Vol. VIII, Springer-Verlag, New York, 1978.
38. D. E. Troxel, et al., "A Two-Channel Picture Coding System: I — Real-Time Implementation," IEEE Trans. on Communications, Vol. COM-29, No. 12, December 1981, pp. 1841-1848.
39. L. G. Roberts, "Picture Coding Using Pseudo-Random Noise," IRE Trans. on Information Theory, Vol. IT-8, February 1962, pp. 145-154.

## Bibliography

E. C. Cartarette, et al., ed., *Handbook of Perception*, Vol. V, "Seeing," Academic Press, 1975.

T. N. Cornsweet, *Visual Perception*, Academic Press, 1970.

*Handbook of Sensory Physiology*, Springer-Verlag, 1973.

D. E. Pearson, *Transmission and Display of Visual Information*, Pentech Press, 1975.

C. G. Mueller, *Sensory Psychology*, Prentice-Hall, 1965.

G. J. Tonge, "Signal Processing for HDTV," IBA Report E8D, July 1983, USA.